用户交互对社会标注行为的差异影响研究*

——以豆瓣网为例

■ 庄倩[1] 骆慧颖[1,2] 戴崇丞[1] 刘丽霞[1] 靳雪宁[1]

[1] 南京农业大学信息科学技术学院 南京 210095 [2]北京师范大学系统科学学院 北京 100875

摘 要：[目的/意义]为提高标签质量，优化社会标注系统的信息服务提供依据，从用户在社会标注系统中与其他用户交互的视角，探讨不同交互特征用户的标注行为差异。[方法/过程]以豆瓣读书作为社会标注系统研究样本，从标签数量、标签结构、标签语义、标注动机和活跃度五个角度研究豆瓣用户标注行为的分布特征；使用用户的关注人数、被关注人数和使用年限表征其在社会标注系统中与其他用户的关联和交互程度，通过差异性分析探讨不同交互特征用户标注行为的差异，并通过多元回归分析研究交互特征对这种差异影响的程度。[结果/结论]实证研究表明，不同交互特征的用户间的社会标注行为存在显著差异：与其他用户交互比较强的用户标签数据集包含更多的标签，关注其他用户越多的用户和被越多用户关注的用户所使用的标签数量越多；使用豆瓣读书年限越长的用户，其平均标签长度和标签重用率越大，而其与其他用户的关注关系对其标签平均长度和标签重用率的影响不大；用户的特殊语种标签比受用户的使用年限影响很大，但一个用户被多少人关注不会显著的影响其特殊语种标签比；关注其他用户越多的用户在标注系统中越活跃。由此可见，社会标注系统可以采取措施加强系统中用户间的交互，通过用户间的相互作用规范用户的社会标注行为，从而提高标签质量。

关键词：用户 交互 标注行为 差异性

分类号：G254

DOI：10.13266/j.issn.0252-3116.2020.20.013

社会标注系统是Web2.0的一个重要应用，是一个描述、组织、管理和分类网络资源的系统。在社会标注系统这个自由开放的网络环境里，用户根据自身的知识水平和认知方式理解网络资源，然后选择自己认知结构中的词语对网络资源添加自然语言（社会标签）。用户可利用社会标注系统上的标签进行信息组织与检索，从而促进信息的传播与共享。由于社会标签的标注采用自由标引方式，部分标签并不能准确地描述资源的内容，用户标注的标签存在着标注随意、无序、标签间缺乏语义关系等质量问题，这在一定程度上影响了基于标签的信息组织与信息检索的效果。标签是用户标注行为的结果，因此自社会标注系统产生以来，用户社会标注行为的研究就引起了学者们的持续关注。在已有研究中，大多选取描述标签的一些量化指标来刻画用户的标注行为，采用统计分析或问卷调查方法，从标注系统中用户整体的标注行为、不同类型用户标注行为等角度展开研究。但是，在社会化标注系统中，用户不但可以基于个人偏好用任意的词语标注网络资源，还可以与有相同兴趣的人传播和分享信息。

社会标注系统本质上是一个全新的、交互的信息空间[1]，因此在对用户社会标注行为的研究中不能只根据用户的客观特征进行分类，还应该考虑用户在社会标注系统中与其他用户的各种交互行为所带来的用户标注行为的差异。深入理解用户标注行为，为优化社会标注系统的设计，提高社会标签质量，改善基于标签的信息组织与信息检索提供一定的理论依据。

* 本文系南京农业大学中央高校基本科研业务费专项资金"多元化视角的用户标注行为及影响因素研究"（项目编号：KYZ201864）和南京农业大学2018年国家大学生创新训练计划项目"豆瓣用户标注行为差异性研究"（项目编号：20181037077）研究成果之一。

作者简介：庄倩（ORCID：0000-0002-0984-4723），讲师，博士，E-mail：zhuangqian@njau.edu.cn；骆慧颖（ORCID：0000-0003-4587-9509），本科生；戴崇丞（0000-0003-2923-9640），本科生；刘丽霞（0000-0002-6310-2157），本科生；靳雪宁（0000-0002-1331-7440）：本科生。

收稿日期：2020-04-13 修回日期：2020-06-17 本文起止页码：117-128 本文责任编辑：杜杏叶

1 相关研究

1.1 用户标注行为的统计特征

理解用户标注行为是社会标签研究领域一个永恒的话题。在对用户标注行为的研究中，大多数对用户标注行为的描述都是从用户的标注结果即标签入手，通过对标签的统计特征来探讨用户的标注行为。例如U. Farooq等从标签增长、标签重用、标签显隐性、标签歧视、标签频率和标注方式来分析用户标签标注行为[2]；V. Mirzaee等从每个资源的标签数量、标签的选择与使用、标注频率等角度分析用户标注动机对标注行为的影响[3]；X. Wang等从用户标签数量、用户标签共享、高频率标签的使用等角度分析了StumbleUpon与Delicious的用户标注行为[4]；A. Guyot从标签长度、标签个数、高频率使用标签、不同语言标签、标签长尾性等角度对LibraryThing中的图书标签进行了分析[5]；S. A. Golder和B. A. Huberman对Delicious系统中用户使用标签的目的、标签的类型以及资源上标签的频率等多方面进行了用户标注模式分析[6]。胡潜等从人均标签个数和标签类型分布两个角度分析了用户标注行为与图书主题之间的关系[7]。

1.2 用户标注行为的差异研究

在用户标注行为的相关研究中，通过对标签数据的统计，分析不同类型用户标注行为差异的研究也有很多。其中，不同知识背景和个人能力的用户会表现出不同的标注行为，研究发现：美国用户添加的标签个数一般显著多于中国和印度用户[8]，而且他们会添加更多表征其主观判断或态度类型的标签[9]；不同职业、专业及职称的科研用户在标注行为上也存在着差异[10]；老用户比新用户更愿意共享标注成果[11]；对学术文本进行标注时专家用户选择的标签与原文的关键词有更高的一致性[12]；没有接受过人工语言或“机读语言”培训的用户在进行标注时更倾向于使用对于目标资源阐释功能更强的复合词标签和多字标签[13]。除此之外，也有研究表明不同的认知特征的用户会表现出不同的标注行为[14-17]，像不同认知难度和认知风格的用户会表现出不同的标签使用行为[18]。通过实验研究还发现不同标注动机的用户也表现出不同的标注行为：以资源分类为动机的用户更倾向于使用固定的标签集合，而且一类资源添加的标签个数基本一致且极少使用同义标签；而以资源描述为动机的用户，标签集合则比较开放，且为每一个资源添加的标签个数经常不稳定，还经常使用同义标签[19]；同时，标注动机也会影响用户的标注强度和标注频率[20]。

1.3 用户标注行为的影响因素研究

在用户标注行为的差异研究中，研究者根据用户的特征将用户划分为不同的群体来研究他们标注行为的差异，本质上就是研究用户自身的因素对社会标注行为的影响。除此之外，社会标注系统和被标注的资源性质等外界环境因素也会影响用户的标注行为[7,21]。

社会化标注的系统因素主要包括社会标签系统中的功能、界面设计、标注规则、其他用户的标签。系统因素对用户标注行为具有很大影响，在这方面已有的研究主要关注系统中的标签推荐和标注规则，以及其他用户标注行为的影响。吴丹等[22]从不同的标签角度将武汉大学图书馆图书标注与豆瓣网图书标注进行对比，发现了图书馆与图书分享网站上用户标注行为的差异。谢佳琳等[23]以武汉大学图书馆用户为研究对象，以信息系统成功模型为框架研究了高校图书馆标注系统质量对其用户标注行为的影响。庄倩等[24]以信息系统成功模型为框架探讨了社会标注系统质量对于用户标注意愿的影响。用户的标注行为除了受系统本身的功能和界面的影响外，基于社会认同理论，系统的标注规则、标签推荐规则、其他用户标注的标签对用户的标注行为起到更大的作用，即社会标签系统中用户于用户、用户与系统间的交互对标注行为都有很大影响。J. B. Philip[25]通过调查分析发现社会认同会影响用户的标注行为；社会化标注系统中的标签推荐和社会化标注系统中其他用户的社会标注行为具有相同的功能，使不同用户的标签具有较高的一致性。T. Kowatsch[26]等人通过实验研究发现50%用户使用了提前加入社会标注系统的预定义词语。C. Dan[27]等在MoviLens推荐系统上的实验研究发现：推荐系统会引导用户的标注行为趋于一致性，用户都会受到系统推荐信息的影响。M. Cameron[28]研究了Flickr用户之间的标签词汇的重叠问题，发现是朋友关系的用户之间比随机抽取的用户之间在标签词汇有更高的重叠度。

资源作为用户的标注对象也是影响标注行为的重要因素，如对于图片、图书、视频等不同的网络资源，用户的标签使用行为有显著差别。Y. Choi等[29]等通过对人文领域数字化档案资料的标签分析发现，非文本资源和文本资源的平均标签数量和标签类型有显著差异。M. Heckner等[30]研究了不同的社会标签系统中用户标注行为，发现用户在标注不同资源时使用的标

签类型不同。胡潜等[7]基于豆瓣网的用户数据进行了实证分析，发现图书主题对用户的标注行为有显著的影响。M. Strohmaier 等[31]通过实证研究发现用户间使用标签的一致性更多地体现在对资源的描述上，而不是在资源的分类上。

综上所述，国内外学者已经选取标签个数和标签长度等量化指标对不同类型的用户的标注行为差异及可能影响用户标注行为的因素进行了探讨。但社会标注系统中用户标注行为的研究仍然是研究热点[32]。用户为资源标注的标签表面上可以在个人信息空间内组织和索引信息；然而，通过对标签隐性的或显性的分享和协作，通过标签标注的信息能够被其他用户浏览和搜索[33-34]，这体现了标签的交流性质。通过创建具有社会意义的标签，能够增强与其他标签用户共享信息和想法的可能性。因此，社会标注系统中用户间的交互是影响用户标注行为的重要因素。但通过上述综述发现，现有的研究大多是通过实验来探讨用户交互对社会标注行为的影响，对用户在社会标注系统中与其他用户的关联和交互程度的不同所产生的标注行为模式和规律上的差异没有更深一步的研究。为了更全面系统地揭示用户标注行为的规律，本文将从用户在社会标注系统中与其他用户的关联和交互程度的角度对用户进行分类，研究不同角度下用户标注行为量化指标的分布特征，探讨不同类型用户社会标注行为的差异性，从而为社会标注系统提供个性化、差异化的用户服务提供理论依据。

2 数据与量化指标设计

为了探讨用户社会标注行为的差异，本文从豆瓣网上采集用户标注信息，根据用户与其他用户的交互信息进行不同角度的用户分类，根据用户的标注信息计算用户标注行为的量化指标，从而探讨不同类型用户标注行为的差异性。

2.1 数据采集

豆瓣网是国内典型的社会标注系统。豆瓣用户可以对豆瓣网上的图书、电影、音乐等资源添加社会标签，同时也可以通过社会标签查找自己感兴趣的资源，既实现了用户对豆瓣网上资源的组织和管理，同时也通过用户标注的标签实现了信息检索。因此，本研究选取豆瓣图书的标注数据作为研究样本，能够较好地代表社会标注系统的特征和属性。

本研究的实验数据均采自于豆瓣网的“豆瓣图书”，首先选定一个用户作为采集的起点，将其标注过的所有图书加入到资源集合中，再获取标注资源集合中每本图书的用户建立用户集合，如此迭代下去，不断获取用户的标注数据。截止到 2017 年 12 月 6 日，通过清洗后共获取用户标注信息 15597 条。在获取的数据集中，用户信息包括：用户 ID、注册时间、关注人数、被关注人数和用户从注册以来标注的所有标签。

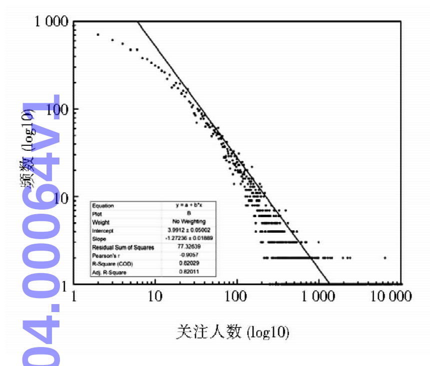
2.2 用户交互特征指标设计及分析

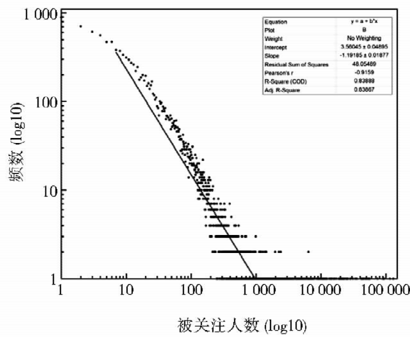
不同的社会标注系统具有不同的社会标注规则和用户标注权限，本研究只关注用户可以自由标引的广义社会标注系统。在这样的在社会标注系统中，允许任意用户对任意资源添加任意数量的非受控自然语言（标签），用户可以选择是否查看某个资源其他用户标注的标签，也可以选择其他用户标注过的标签作为自己的标签，像 Delicious、Flicker、豆瓣网都是这样的社会标签系统。由此，在社会标注系统中用户之间通过标注的标签和资源达成了信息交换，形成了一定程度的在线社会化交互[35]。在像知乎、微博这类社交媒体中，用户间的社会化交互可以通过各种各样的社交互动功能实现，如发布文章、评论、转发、收藏、点赞、关注、分享等[36-37]。但在社会标注系统，用户间的这种关联和交互很难被捕捉和记录。用户的标注行为过程可简化为用户通过网络浏览或口碑传播等方式获得某一资源，用户将该资源存入自己合适的集合中并用相关主题的标签标注该资源两个步骤。在获取标注资源的过程中，用户间的交互就可能已经发挥了作用，除了偶遇外，可能由于其他用户的标注或共享使得用户发现自己感兴趣的资源，同时也获取了其他用户为该资源标注的标签。在标注资源时，用户可以参考其他用户标注的标签。因此，在社会标注行为的整个阶段，用户间的交互都可能影响用户的标注行为却无法被系统记录下来，也就无法获取用户间是否进行交互的数据。这也是在现有关于用户交互对社会标注行为的影响研究中，大多使用实验法进行研究的原因。但实验法也存在一些问题，社会化标注系统本来以用户为中心，可以充分体现用户的自主性和个性化的系统。通过实验很难获取用户真实的标注动机和标注行为。

用户自定义的标签一定程度上反映了用户的行为和偏好，用户根据标签能很容易的找到熟人、合作者，最终和有相同兴趣的人形成社区或找到自己感兴趣的资源。豆瓣网提供了用户间“关注”的功能，用户可以在标注过程中发现自己感兴趣的其他用户并关注，同理一个用户也可能引起其他用户的兴趣而被关注。关注用户更容易获取被关注用户关注的资源、标注的标

签等信息,一定程度上反映了用户之间的交互。这种相互关注的关系也会受到注册时间的影响,因为标注行为是一个持续且漫长的过程,发现志同道合的“朋友”和被别人发现都需要一定的时间,因此用户使用社会标注系统的时间也一定程度的反映了用户间交互的深浅。基于此,本研究通过用户关注其他用户的人数、被关注的人数和使用年限三个指标来表征用户间的关联和交互特征。

图 1 分别给出了用户关注人数、被关注人数和使用年限的统计分布图,其中用户关注人数和被关注人数的横纵坐标分别取了对数。从图中可以看出,用户关注人数和被关注人数的分布均呈现幂律分布,关注人数和被关注人数较多的用户占比较少,大多数用户关注的用户和关注他的用户都比较少;而使用年限的分布基本服从正态分布。为了更清晰的比较关注人数、被关注人数和使用年限这三个用户交互特征对用户标注行为的影响,本研究将用户关注人数和被关注人数由高到低排列后,取前 25% 代表交互程度较高的组,取后 25% 作为交互程度较低的组。而使用年限以众数 8 年作为分界,分为高使用年限组和低使用年限组。





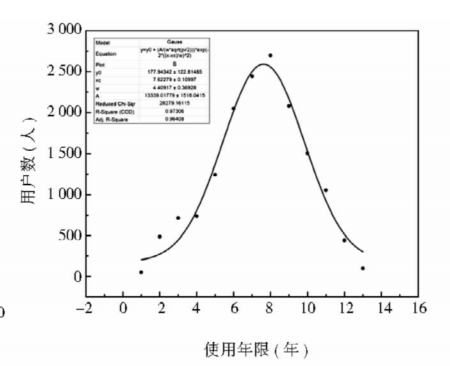


图 1　用户关注人数、被关注人数和使用年限的统计分布

2.3　用户标注行为量化指标设计

在对用户标注行为的现有研究中,描述用户标注行为的量化指标都是标签的统计特征,主要表现为标签的结构和语义两个角度。其中,较多的研究用标签标注频率、标签重用率、标签数量等指标。基于已有的研究和豆瓣用户使用标签的特点,为了从多个角度量化用户的标注行为,本研究选取从数量角度描述用户标注行为的标签个数指标,从结构角度描述用户标注行为的平均标签长度和特殊语种标签比指标,从语义角度描述用户标注行为的标签重用率指标(同时,标签的重用率也会在一定程度上反应用户的标注动机[38]),以及描述用户标注行为强度的指标用户活跃度。每个指标的具体说明如下:

(1)标签个数。用户标签个数是社会标注行为研究中比较常用的量化指标。它指的是一个用户从注册到采集数据时为止所使用的标签总数。用户的标签个数越多,说明用户的标注积极性越高。

(2)平均标签长度。用户平均标签长度指的是每个用户所使用标签的平均字符数,也是比较常用的一个表征用户标注行为的量化指标。为了方便计算,本文在计算标签长度时将 1 个汉字算为 1 个字符,1 个英文字母也算为 1 个字符。该指标反映了用户在标注过程中对所标注词语长度的偏好。

(3)特殊语种标签比。直观来看,一般用户选择标签的语种都与社会标注系统或资源的语种一致。但在分析用户的标签时我们发现,有些豆瓣网用户使用一定比例的非汉语标签。这些标签大多反应的是用户自身对资源的理解或情感,不是客观描述资源的词语,是比较个性化的标签。因此,我们提出了特殊语种标签比这个量化指标来描述标签的结构。我们将特殊语种标签定义为含有非汉语字符的标签。特殊语种标签比表示用户使用特殊语种标签的比例,在一定程度上反映了用户对个性化标签的偏好。

(4)标签重用率。标签重用率表示用户使用每个标签的频次之和与不同的标签总数之比,因此该值大于或等于 1。标签的重用率不但反映了用户标签的语义,也一定程度反映了用户的标注动机,是比较常用的描述标签语义的量化指标。以分类为动机的用户更倾向于使用固定的标签集合描述资源,对一类资源添加的标签个数基本一致,标签的重用率较高;而以描述为动机的用户,为每一个资源添加的标签个数经常不稳定,重用率较低[11]。

(5)用户活跃度。尽管标签数量能够一定程度地描述用户的标注积极性,但考虑到用户的使用年限对

用户标签数量的影响，本文通过用户的年平均标注标签数描述用户的标注强度。用户的活跃程度越高，说明用户一年内标注的标签越多，标注强度越高。

上述用户标注行为量化指标计算方法说明如表 1 所示：

表 1　用户标注行为量化指标及说明

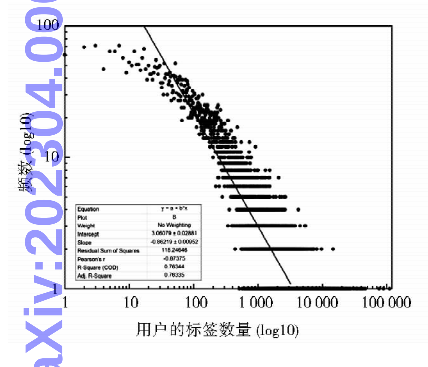
指标选取角度	用户标注行为量化指标	指标公式
标签数量	标签个数	Σ不去重的标签总数
标签结构	平均标签长度	(Σ标签的字符数)/标签总数
	特殊语种标签比	(Σ含非汉语字符的标签个数)/标签总数
标签语义	标签重用率	(Σ标签的使用频次)/(去重后的标签总数)
标注强度	用户活跃度	标签总数/使用年限

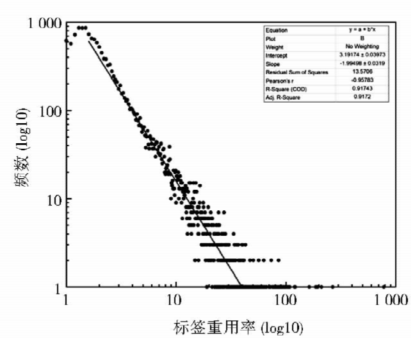
3　结果分析

3.1　用户标注行为量化指标分析

社会标注系统是一个大众参与的复杂系统，而幂律分布能够更真实地反映复杂系统的复杂性。通常，将幂律分布分为头部和尾部，头部包含少量的使用频率较高的元素，尾部包含大量的使用频率较低的元素，通常称为“长尾”[39]。本研究对上述五个量化用户标注行为的指标进行了统计分析。

图 2 给出了双对数坐标下用户标签个数、标签重用率和用户活跃度的分布图。从图中可以看出，这三种量化指标的分布均服从“长尾”分布。说明与 Delicious，Flickr 等典型的社会标注系统一样，豆瓣网上少部分用户使用了大量的标签，大部分的用户使用了少量的标签；豆瓣用户在使用标签时有大部分用户的标签重用率较低，只有较少的用户重复使用固定的一些标签，这一定程度上说明大部分豆瓣用户的标注动机是描述资源，而不是组织和分类资源；豆瓣用户中有大部分用户的标注强度较低，少量的用户标注强度较高，说明豆瓣的注册用户中活跃度高的用户较少。





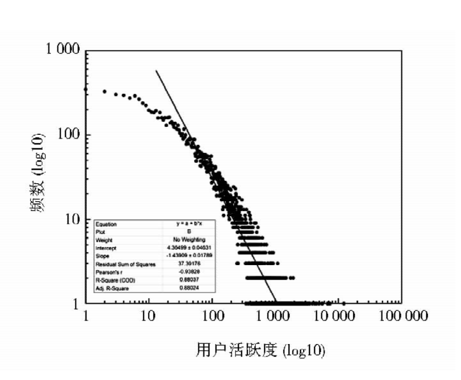


图 2　用户标签个数、标签重用率、用户活跃度的统计分布

将用户使用的标签平均长度作为横坐标，平均标签长度为该标签长度的用户频次取对数后作为纵坐标，得到了如图 3 所示的分布图。从图中可以看出，使用平均长度为 1 的标签的用户很少，而使用标签平均长度为 2、3、4 的用户约占所有用户的 94%，几乎 61% 的用户的平均标签长度为 3。这主要是受到汉语语言习惯的影响，因为汉语中很少有一个字的词语，大多数词语都由两个汉字或四个汉字组成，这就使得用户的平均标签长度集中于 2、3、4 个汉字。平均标签长度在 5 以上的用户仅占 0.06%，说明有很少的用户用词组或句子去标注资源。这个统计结果反映出用户标注行为在一定程度上符合最小省力法则，即用户更倾向于选择用尽可能少的汉字标注资源。

本研究中提出量化标签结构的指标特殊语种标签比来描述用户的标签集合中非汉语标签所占的比例。经统计发现，特殊语种标签比为 0 的用户数量并不多，

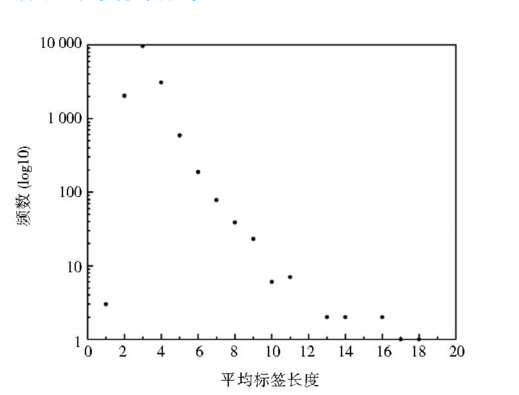


图 3　用户平均标签长度统计分布

这表明绝大多数豆瓣用户在标注时都使用了汉语之外的标签。图 4 给出了特殊语种标签比的分布图，其尾部基本符合“长尾”特征，即使用非汉语标签较多的用户相对较少。从图中可以看出，绝大多数用户的特殊语种标签比分布在 0 到 0.1 范围内，这说明多数用户使用的标签中，特殊语种标签占比重较小，豆瓣用户更

倾向于使用汉语标签,很少使用其他语言、字符和数字等,而其他语言或符号的标签一般是用户用来描述资源、表达情感的个性化标签,这个结果从形式上说明了个性化的标签占比较少。但还是有极少数的用户的特殊语种标签比为 1,即有用户使用了完全非汉语的个性化标签,这些个性化标签的存在也是造成社会标签质量较差[40]的原因之一。

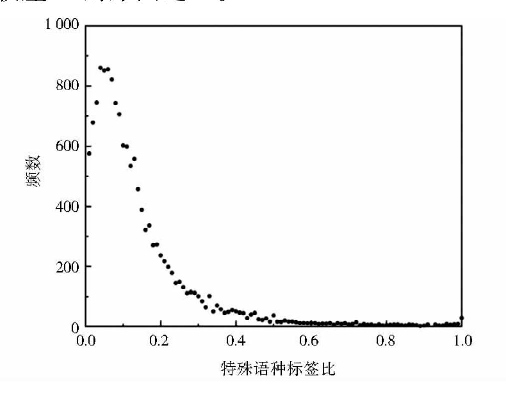


图 4 用户特殊语种标签比分布特征

在已有研究中通过对不同社会标注系统中标注数据的统计分析,发现社会标注系统中标签数量的增长[41-42]、用户使用标签的分布[43-45]、资源被标注标签的分布、标签使用频次的分布[46]、标签重用率的分布都服从幂律分布。但现有研究对豆瓣网标注数据的统计分析很少或几乎没有。上述五个量化指标的统计分析结果表明:豆瓣网用户的标签个数、标签重用率和用户活跃度的分布都服从幂律分布,豆瓣网的用户标注行为与其他社会标注系统的用户标注行为具有一致的宏观特征,这进一步说明了豆瓣网可以作为研究用户标注行为的平台;豆瓣网用户在进行标注时偏好使用字数为 2-4 个汉字的词语作为标签,基本符合汉语的语言特征。本文提出的量化用户使用个性化标签程度的指标特殊标签比的统计分布可以看出,绝大部分的豆瓣用户都会使用个性化的标签,但是在标签集中占比不高,说明豆瓣网用户还是倾向于使用更大众化的标签。

3.2 用户交互特征与标注行为关系分析

为了验证用户的交互特征是否回带来用户标注行为的差异,本节对用户交互特征中的高低关注人数、被关注人数和使用年限等二分类别变量进行独立样本检验。由于各类别中数据不满足方差齐性的条件,所以利用 SPSS 26 进行两独立样本 Mann-Whitney U 检验,探讨用户间不同交互程度对用户标注行为的影响,结果如表 2 所示:

表 2 不同交互特征用户标注行为的差异比较

变量	类型	描述统计量	标签个数	平均标签长度	标签重用率(%)	特殊语种标签比(%)	标注强度
关注人数	低(N=3900)	均值	495.37	3.12	3.32	0.11	79.07
		中位数	154	2.95	1.76	0.068	26.57
		方差	1 141.22	0.94	5.36	0.13	178.76
	高(N=3900)	均值	1 939.45	3.35	5.07	0.15	231.37
		中位数	768	3.19	2.36	0.11	94.36
		方差	4 035.48	0.91	9.89	0.15	480.84
	Z 值	——	-37.288	-16.169	-20.066	-20.783	-31.768
	Sig.	——	.000	.000	.000	.000	.000
被关注人数	低(N=3900)	均值	363.1	3.08	2.76	0.10	71.09
		中位数	122	2.94	1.65	0.06	24.36
		方差	814.44	0.85	4.44	0.13	168.23
	高(N=3900)	均值	2 457.91	3.42	6.71	0.18	269.8
		中位数	1167	3.24	2.93	0.13	127.67
		方差	4 526.332	0.99	12.45	0.17	496.97
	Z 值	——	-50.666	-20.033	-35.879	-29.997	-40.994
	Sig.	——	.000	.000	.000	.000	.000
使用年限	低(N=7725)	均值	689.66	3.13	3.49	0.11	126.35
		中位数	211	2.99	1.82	0.07	41.17
		方差	1 759.71	0.81	10.63	0.13	306.81
	高(N=5178)	均值	1 868.48	3.38	5.82	0.17	184.12
		中位数	802	3.2	2.61	0.12	81.5
		方差	3 786.80	0.96	10.84	0.16	369.41
	Z 值	——	-40.344	-17.913	-30.006	-25.505	-20.888
	Sig.	——	.000	.000	.000	.000	.000

表 2 的检验结果显示：高低关注人数的用户组在表征用户标注行为的各个维度上都有显著性差异(Sig. <0.05)。从均值上看，高关注人数用户组的标签个数、平均标签长度、标签重用率、特殊语种标签比和标注强度均高于低关注人数用户组且相差较大。这一定程度上是由于高低关注人数的定义选取的是前后各 25% 的数据，包含了极端值。为了进一步说明高低关注人数组在用户标注行为的差异，表 2 中给出了各组各维度下的中位数。虽然中位数之间的差值远远小于均值之间的差异，但五个维度下，高关注人数组还是均高于低关注人数组。

由表 2 还可知：不同的被关注人数在用户标注行为的五个维度上均存在显著性差异(Sig. <0.05)。而且，同关注人数一样，高被关注人数组的标签个数、平均标签长度、标签重用率、特殊语种标签比和标注强度无论均值还是中位数均高于低被关注人数组。

表 2 中还给出了不同使用年限用户组用户标注行为的差异分析结果，同关注人数和被关注人数一样，高低使用年限的用户组在标签个数、平均标签长度、标签重用率、特殊语种标签比和标注强度均存在显著性差异(Sig. <0.05)，且高使用年限组的用户标注行为的五个维度显著高于低使用年限用户组。

通过以上的差异性分析可以看出，不同交互特征的用户在标注行为的五个量化指标上均存在显著性差异。说明用户间的交互会在一定程度上影响用户的标注行为，是用户标注行为的影响因素。

3.3 用户标注行为多元回归分析

为了进一步探讨不同交互特征对用户交互行为的影响强弱程度。本小节对用户标注行为进行多元回归分析。为了更精确的研究用户交互特征与用户标注行为之间的关系，在多元回归分析中使用了所有 15 597 名用户的完整数据集。

3.3.1 标签数量的多元回归分析

以标签数量(TN)为因变量，以关注人数(CN)、被关注人数(CDN)和使用年限(UY)作为自变量，采用逐步多元回归法进行分析。结果显示，关注人数(CN)、被关注人数(CDN)和使用年限(UY)均被纳入回归方程。选择最优的回归模型，回归系数及显著性检验结果如表 3 所示。其中，容忍度和方差膨胀系数(VIF)值可以检验多元回归分析是否有多元共线性问题：容忍度值越接近 0，表明变量间的多重共线性问题越严重；VIF 值如果大于 10，则表明变量间有多重共线性问题[47]。本回归模型的容忍度值都大于 0.94，VIF 值小于 1.06，表明不存在多重共线性问题。下面对标签平均长度、标签重用率、特殊语种标签比和标注强度的多元回归分析的自变量均为关注人数、被关注人数和使用年限，故都不存在多元共线性问题，下面的分析中不再赘述。通过多元回归分析，得到如下回归方程：

$$TN = 0.178\ CN + 0.171\ CDN + 0.045\ UY \quad 方程(1)$$

表 3 用户标签数量的多元回归系数分析表

自变量	未标准化系数(B)	标准误(SE)	标准化系数(β)	t	Sig.	容忍度	VIF
常量	-609.663	66.479		-9.171	.000		
关注人数	2.249	0.100	0.178	22.569	.000	0.949	1.054
注册时间	189.650	8.717	0.171	21.757	.000	0.957	1.045
被关注人数	0.048	0.008	0.045	5.750	.000	0.981	1.019

此外，通过表 4 可知，F(3,15593) = 441.824，P = 0.000，达到小于 0.001 的显著性水平，表明回归显著，说明了该回归方程的有效性。

表 4 用户标签数量的多元回归方差分析表

模型	平方和	自由度	均方	F	显著性
回归	8 940 693 781.967	3	2 980 231 260.656	441.824	.000
残差	105 179 333 428.046	15 593	6 745 291.697		
总计	114 120 027 210.012	15 596			

3.3.2 平均标签长度的多元回归分析

以平均标签长度(TL)为因变量，以关注人数(CN)、被关注人数(CDN)和使用年限(UY)作为自变量，采用逐步多元回归法进行分析。通过表 5 可知，关注人数(CN)、被关注人数(CDN)和使用年限(UY)均被纳入回归方程。通过表 6 可知，F(3,15593) = 79.73，P = 0.000，达到小于 0.001 的显著性水平，表明回归显著。回归方程如下：

$$TL = 0.044\ CN + 0.017\ CDN + 0.103\ UY \quad 方程(2)$$

表 5 用户平均标签长度的多元回归系数分析

自变量	未标准化系数(B)	标准误(SE)	标准化系数(β)	t	Sig.
常量	2.912	0.023		124.026	.000
关注人数	0.000	0.000	0.044	5.374	.000
注册时间	0.039	0.03	0.103	12.691	.000
被关注人数	6.089E-6	0.000	0.017	2.084	.037

表 6 用户平均标签长度的多元回归方差分析

模型	平方和	自由度	均方	F	显著性
回归	201.29	3	67.1	79.73	.000
残差	13 122.23	15 593	0.842		
总计	13 323.522	15 596			

3.3.3 标签重用率的多元回归分析

以标签重用率(TR)为因变量,以关注人数(CN)、被关注人数(CDN)和使用年限(UY)作为自变量,采用逐步多元回归法进行分析。回归结果显示,关注人数(CN)、被关注人数(CDN)和使用年限(UY)均被纳入回归方程。选择最优的回归模型,回归系数及显著性检验结果如表 7 所示,回归系数的显著性检验结果均小于 0.05,通过检验。通过表 8 可知,F(3,15593) = 79.73,P = 0.000,达到小于 0.001 的显著性水平,回归显著。多元回归法得到如下回归方程:

$$TR = 0.023\ CN + 0.027\ CDN + 0.104\ UY \quad 方程(3)$$

表 7 用户标签重用率的多元回归系数分析

自变量	未标准化系数(B)	标准误(SE)	标准化系数(β)	t	Sig.
常量	0.964	0.265		3.632	.000
关注人数	0.001	0.000	0.023	2.759	.006
注册时间	0.447	0.035	0.104	12.829	.000
被关注人数	0.000	0.000	0.027	3.329	.001

表 8 用户标签重用率的多元回归方差分析

模型	平方和	自由度	均方	F	显著性
回归	23 226.19	3	7742.06	71.97	.000
残差	1 677 426.85	15593	107.58		
总计	1 700 653.03	15596			

3.3.4 特殊语种标签比的多元回归分析

以特殊语种标签比(TC)为因变量,以关注人数(CN)、被关注人数(CDN)和使用年限(UY)作为自变量,采用逐步多元回归法进行分析。回归结果显示,被关注人数(CDN)被排除,关注人数(CN)和使用年限(UY)被纳入回归方程。选择最优的回归模型,具体回归方程的回归系数及显著性检验结果见表 9。表 9 展示了特殊语种标签比回归模型的回归系数及显著性检验结果,显著性 Sig. 均小于 0.01,表明模型通过显著性检验。此外,上述回归模型的容忍度值为 0.96,VIF 值为 1.041,表明不存在多元共线性问题。通过表 10 可知,F(2,15594) = 272.137,P = 0.000,达到小于 0.001 的显著性水平,表明回归显著。利用回归分析法得到回归方程为:

$$TC = 0.039\ CN + 0.172\ UY \quad 方程(4)$$

3.3.5 用户标注强度的多元回归分析

以用户标注强度(TA)为因变量,以关注人数(CN)、被关注人数(CDN)和使用年限(UY)作为自变量,采用逐步多元回归法进行分析。结果显示,关注人数(CN)、被关注人数(CDN)和使用年限(UY)均被纳入回归方程。选择最优的回归模型,回归系数及显著性检验结果如表 11 所示。回归系数的显著性检验结果 Sig. 均小于 0.01,通过显著性检验。得到如下回归方程:

$$TA = 0.185\ CN + 0.04\ CDN + 0.032\ UY \quad 方程(5)$$

表 9 用户特殊语种标签比的多元回归系数分析

自变量	未标准化系数(B)	标准误(SE)	标准化系数(β)	t	Sig.	容忍度	VIF
常量	0.053	0.004		14.042	.000		
关注人数	0.000027	0.000	0.039	4.867	.000	0.96	1.041
注册时间	0.010	0.000	0.172	21.389	.000	0.96	1.041

表 10 用户特殊语种标签比的多元回归方差分析

模型	平方和	自由度	均方	F	显著性
回归	11.67	2	5.835	272.137	.000
残差	334.357	15 594	0.021		
总计	346.027	15 596			

表 11 用户标注强度的多元回归系数分析

自变量	未标准化系数(B)	标准误(SE)	标准化系数(β)	t	Sig.
常量	64.580	8.094		7.979	.000
关注人数	0.278	0.012	0.185	22.924	.000
注册时间	5.245	1.061	0.040	4.942	.000
被关注人数	0.004	0.001	0.032	4.072	.000

此外,通过表 12 可知,F(3,15593) = 223.63,P = 0.000,达到小于 0.001 的显著性水平,表明回归显著,说明了该回归方程是有效的。

表 12 用户标注强度的多元回归方差分析

模型	平方和	自由度	均方	F	显著性
回归	67 084 225.32	3	22 361 408.44	223.63	.000
残差	1 559 163 325.12	15 593	99 991.23		
总计	1 626 247 550.44	15 596			

4 研究结果讨论

4.1 用户标注行为的个体差异性

由 3.2 节的结果分析可以看出，用户的关注人数、被关注人数和使用年限都是造成用户标注行为差异的因素。对于关注人数、被关注人数和使用年限这三个用户交互特征而言，高关注人数、被关注人数用户组在用户标注行为的五个维度上均高于低关注人数、低被关注人数用户组，使用年限长的用户组在用户标注行为的五个维度上均高于使用年限短的用户组。也就是说，与其他用户交互强度高的用户与交互强度弱的用户在标注行为上是有显著差异的。产生这样结果的原因可能是，社会标注系统是一个自由开放的网络环境，其所包含的内容和功能十分广泛，且用户的标注动机和标注行为都是不受控的。如果用户要利用它进行网络信息资源的分类、组织和检索，需要用户经过较长时间的接触、尝试和学习过程。而与其他用户交互较多的用户可以更容易、更快速的完成这一过程。较高的标签重用率即体现了这些用户的标注动机不仅仅是描述，而上升到了利用标签为自己感兴趣的资源分类。较高的平均标签长度和特殊语种标签比说明这些用户已经熟练地掌握了标注过程，开始标注具有个人特征的个性化标签。同时，对社会标注系统标注功能和标注过程的熟练能够提高用户对于标注系统的满意度和兴趣，从而促使用户有更高的活跃度，标注更多的资源，也因此具有了更多的标签数量。

然而，这个结果也在一定程度上说明了社会标注系统的界面设计和导航功能有待改善。如果其界面设计的提示和导航功能比较完善，那么用户就不需要花费更多的时间尝试和学习标注过程，即使是低交互强度的用户也会有和高交互强度用户同样的标注行为。

4.2 用户交互特征对用户标注行为的影响

对不同交互特征的用户标注行为的差异性研究发现，用户的交互特征会影响用户的标注行为。在已有的用户标注行为影响因素的研究中并没有考察用户交互的影响。本研究通过进一步的多元回归分析得到了初步的结论。将用户的交互特征作为自变量，用户标注行为的每个维度作为因变量，利用逐步多元回归法进行分析，形成最优回归方程。研究结果显示：用户的关注人数、被关注人数和使用年限都会对用户的标签数量、平均标签长度、标签重用率和标注强度产生影响；但特殊语种标签比只受到用户关注人数和使用年限的影响，被关注人数的多少并不能影响用户的特殊语种标签比。通过回归方程的系数发现：用户的关注人数、被关注人数和使用年限对用户标签数量的影响都是正向的，影响系数分别为 0.178、0.171 和 0.045，其中使用年限对标签数量的影响最小，这是因为文中的使用年限是从用户注册豆瓣网的时间开始计算的，说明并不是越早注册的用户标注的标签就越多，可能有些用户标注过后对标注系统并不满意就不再进行标注了。用户的关注人数、被关注人数和使用年限对用户的平均标签长度的影响也都是正向的，分别为 0.044、0.017 和 0.103，影响系数都比较小。用户的关注人数、被关注人数和使用年限对用户的标签重用率的影响系数分别为 0.023、0.027 和 0.104，也都比较小，相对影响最大的是使用年限。可能是由于随着使用年限的增加，用户对社会标注系统的逐步了解使得用户开始尝试利用标签进行资源的分类和组织。三个用户交互特征对用户标注强度影响系数分别为0.185、0.04 和 0.032，影响系数最高的是关注人数。用户在标注系统中关注了更多的其他用户，就有更大的概率从别的用户那里发现自己感兴趣的资源，从而进行标注，提高了自身的标注强度。用户的特殊语种标签比只受到关注人数和使用年限的影响，影响系数分别为 0.039 和 0.172。特殊语种标签比在一定程度上反映了用户在选择标签时是否使用带有强烈个人色彩的个性化标签。产生这样的结果可能是由于在标注的初始阶段用户处于尝试和学习标注的阶段，更容易受其他用户和社会标注系统推荐标签的影响，选用一些比较大众化的标签，这也可以解释用户的关注人数会影响用户的特殊语种标签比的原因。但随着标注行为的熟练，用户开始标注一些表达自己的情感有自身特色的个性化标签，特殊语种标签比提高。

5 结论与展望

为了更全面系统地揭示用户标注行为的规律，为社会标注系统提供个性化、差异化的用户服务提供理论依据，本文以豆瓣网作为社会标注系统的研究平台，从用户交互的角度对用户进行分类，研究用户在社会标注系统中与其他用户的不同关联和交互程度下标注行为的统计特征，探讨不同交互特征用户社会标注行为的差异性。实证研究发现，不同关注人数、被关注人

数和使用年限的用户在标注行为的各个维度上都存在差异,具体结论如下:

(1)豆瓣网用户的社会标注行为统计特征与针对Delicious,Flickr 等典型的社会标注系统已有的研究结论基本吻合,说明豆瓣网可以作为研究用户社会标注行为的平台。

(2)不同交互特征的用户间的社会标注行为存在显著差异:与其他用户交互比较强的用户标签数据集包含更多的标签,关注其他用户越多的用户和被越多用户关注的用户所使用的标签数量越多;使用豆瓣读书年限越长的用户,其平均标签长度和标签重用率越大,而他与其他用户的关注关系对其标签平均长度和标签重用率的影响不大;用户的特殊语种标签比也受用户的使用年限影响很大,但一个用户被多少人关注不会显著的影响他的特殊语种标签比;关注其他用户越多的用户在标注系统中越活跃。

本文的研究结论具有一定的理论意义和实践价值。首先,已有关于交互对用户标注行为影响的研究没有区分用户和用户间的交互还是用户与平台的交互,且大多通过实验法进行研究,实验数据无法完全反应用户真实的标注行为。本文使用大量的真实用户交互和标注数据从定量的角度验证了用户间的交互对用户标注行为的影响进行了验证。其次,通过研究结果,可以为社会标注系统的优化提供参考建议:改进系统"关注"功能的导航和界面,增加用户间的"朋友"关系,可以在一定程度上增加用户的活跃度和标注意愿;通过增加功能提高新老用户之间的交互,从而提高标签的重用率,可以在一定程度上降低个性化标签,进一步规范标签的使用,从而提高信息检索和信息组织等信息服务质量。

虽然本研究通过实证研究已经得到了用户交互特征对用户标注行为有一定影响的结论,但仍然存在一定的局限性。首先,本研究只采集了豆瓣图书这一单一社会标注系统的一部分用户的数据,后续可扩大数据来源和数据规模以便做进一步研究。其次,本研究量化用户标注行为的维度是在已有研究基础上添加了表征用户选择标签特征的特殊语种标签比,但在这些维度中并不包含标签的具体语义,因此下一步研究可以从标签语义的角度来量化用户标注行为。再次,只用了关注人数、被关注人数和使用年限这些间接的静态指标来表述用户之间的交互特征,从时间演化的角度,用户之间的交互会如何影响用户的标注行为还有待进一步的研究。

参考文献:

[1] TRANT J. Studying social tagging and folksonomy: a review and framework [J]. Journal of digital information, 2009, 10(1): 1-42.

[2] FAROOQ U, KANNAMPALLIL TG, SONG Y, et al. Evaluating tagging behavior in social bookmarking systems: metrics and design heuristics [C] //Proceedings of the 2007 international ACM conference on supporting group work. New York: Association for Computing Machinery, 2007: 351-360.

[3] MIRZAEE V, IVERSON L. Tagging: Behaviour and motivations [J]. Proceedings of the American Society for Information Science and Technology, 2009(46): 1-5.

[4] WANG X, KUMAR S, LIU H. A study of tagging behavior across social media [C] //Proceeding of the 2011 workshop on social web search and mining. Beijing: SIGIR, 2011.

[5] GUYOT A. Understanding booksonomies-How and why are book taggers tagging [D]. Aberystwyth: University of Wales, 2013.

[6] GOLDER S A, HUBERMAN B A. Usage patterns of collaborative tagging systems [J]. Journal of information science, 2006, 32(2):198-208.

[7] 胡潜,石宇. 图书主题对用户标签使用行为影响研究[J]. 图书情报工作,2016,60(8):106-112.

[8] PEESAPATI S T, WANG H C, COSLEY D. Intercultural human photo encounters: how cultural similarity affects perceiving and tagging photographs [C]// Proceedings of the 3rd international conference on intercultural collaboration. New York: Association for Computing Machinery, 2010: 203-206.

[9] DONG Z, SHI C, SEN S, et al. War * vs in forrest gump: cultural effects in tagging communities [C]//Proceedings of the sixth international AAAI conference on weblogs and social media. California: The AAAI Press, 2012: 82-89.

[10] 张颖怡, 章成志, 池雪花, 等. 科研用户博文关键词标注行为差异研究——以科学网博客为例[J]. 现代图书情报技术, 2015(10): 13-21.

[11] SEN S, LAM S K, et al, Tagging, community, vocabulary, evolution [C]//Proceedings of the 2006 ACM conference on computer supported cooperative work. New York: Association for Computing Machinery, 2006: 181-190.

[12] ŠPIRANEC S, IVANJKO T. Experts vs. novices tagging behavior: an exploratory analysis [J]. Procedia - Social and Behavioral Sciences, 2013, 73: 456-459.

[13] 查先进,吕彬. 知识共享视角下的大众标注行为研究-基于标签的实证分析[J]. 图书馆论坛, 2010, 30(06): 76-81.

[14] SINHA R. A cognitive analysis of tagging [EB/OL]. [2020-02-25]. http://rashmisinha. com/2005 /09/27/a-cognitive-analy-

sis-of-tagging/.
[15] SZEKELY B, TORRES E. Ranking bookmarks and bistros: intelligent community and folksonomy development [EB/OL]. [2020－02－23]. http://www. eliastorres. com/archives /2005 /07 /13 /tagrank. pdf.
[16] 易明,曹高辉,毛进,等. 基于 Tag 的知识主题网络构建与 Web 知识推送研究 [J]. 中国图书馆学报, 2011, 37(4): 4－12.
[17] 马费成,张斌. 图书标注环境下用户的认知特征[J]. 中国图书馆学报, 2014, 40(1): 4－14.
[18] 林鑫,周知. 用户认知对标签使用行为的影响分析－基于电影社会化标注数据的实证分析 [J]. 情报理论与实践, 2015, 38(10): 85－88.
[19] STROHMAIER M, KÕRNER C, KERN R. Understanding why users tag: a survey of tagging motivation literature and results from an empirical study [J]. Web semantics science services & agents on the World Wide Web, 2012, 17(4): 1－11.
[20] 王娜,马云飞. 网络环境下大众标注行为动机的调查与分析 [J]. 图书情报工作, 2013, 57(23): 100－107.
[21] 常唯. 论网络环境下用户标注的价值与应用 [J]. 图书情报工作, 2008, 52(1): 9－12.
[22] 吴丹, 许小梅. 图书馆与图书分享网站的用户标注行为比较研究[J]. 图书情报知识, 2013(1):85－93.
[23] 谢佳琳, 张晋朝. 高校图书馆用户标注行为研究——以信息系统成功模型为视角[J]. 图书馆论坛, 2014, 34(11):87－93.
[24] 庄倩, 韩正彪. 社会标注系统质量对用户标注意愿的影响机理 [J]. 图书馆论坛, 2019, 39(6): 71－79.
[25] BINKOWSKI P J. The effect of social proof on tag selection in social bookmarking applications [EB/OL]. [2020－03－27]. http://etd. ils. unc. edu:8080/dspace/bit stream/1901/358/1/philipbinkowski. pdf.
[26] KOWATSCH T, MAASS W. The impact of pre-defined terms on the vocabulary of collaborative indexing systems [C] //16th European conference on information systems. Galway: ECIS Standing Committee, 2008: 2136－2147.
[27] COSLEY D, LAM S K, ALBERT I, et al. Is seeing believing? how recommender interfaces affect users' opinions[C]// Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: Association for Computing Machinery, 2003: 585－592.
[28] CAMERON M. Position paper, tagging, taxonomy, flickr, article, toread [EB/OL]. [2020－04－20]. http://ci teseerx. ist. psu. edu/view doc/summary? doi = 10. 1. 1. 74. 8883.
[29] CHOI Y, SYN S Y. Characteristics of tagging behavior in digitized humanities online collection [J]. Journal of the association for information science and technology, 2015, 27(5):1－36.
[30] HECKNER M, NEUBAUER T, WOLFF C. Tree, funny, to read, google: what are tags supposed to achieve? a comparative analysis of user keywords for different digital resource types[C]// Proceedings of the 2008 ACM workshop on search in social media. New York: Association for Computing Machinery, 2008:3－10.
[31] STROHMAIER M, KÖRNER C, KERN R. Understanding why users tag: a survey of tagging motivation literature and results from an empirical study [J]. Web semantics: science, services and agents on the World Wide Web. 2012, 17(12): 1－11.
[32] XU C, MA B J, CHEN X H, et al. Social tagging in the scholarly world [J]. Journal of the American Society for Information Science and Technology, 2013, 64(10): 2045－2057.
[33] LEE B Z, GE S L. Personalisation and sociability of open knowledge management based on social tagging [J]. Online information review, 2010, 34(4): 618－625.
[34] MENDES L H, QUIÑONEZ-SKINNER J, SKAGGS D. Subjecting the catalog to tagging [J]. Library Hi Tech, 2009, 27(1): 30－41.
[35] CHEN C, LIN Y. What drives live-stream usage intention? The perspectives of flow, entertainment, social interaction, and endorsement [J]. Telematics and informatics, 2018, 35(1): 293－303.
[36] ALLEN S M, CHORLEY M, COLOMBO G B, et al. Exploiting user interest similarity and social links for micro-blog forwarding in mobile opportunistic networks[J]. Pervasive and mobile computing, 2014, 11: 106－131.
[37] 邓胜利, 蒋雨婷. 用户交互特征对知识付费行为预测的贡献度研究[J/OL]. 图书情报工作:1－10[2020－06－15]. https://doi. org/10. 13266/j. issn. 0252－3116. 2020. 08. 011.
[38] STROHMAIER M, KÖRNER C, KERN R. Understanding why users tag: a survey of tagging motivation literature and results from an empirical study[J]. Web semantics science services & agents on the World Wide Web,2012,17(4) : 1－11.
[39] 吴振宇,胡军,李德毅. 社会标注系统幂律特性分析[J]. 复杂系统与复杂性科,2014,11(2):5－16.
[40] 李蕾, 王冕, 章成志. 区分标签类型的社会化标签质量测评研究[J]. 图书情报工作, 2013, 57(23): 11－16.
[41] LI D, DING Y, QIN Z, et al. Dynamic features of social tagging vocabulary: Delicious, Flickr and YouTube[C]// 2010 International Conference on Advances in Social Networks Analysis and Mining. New York: Institute of Electrical and Electronics Engineers, 2010: 316－320.
[42] VANDER WAL T. Explaining and showing broad and narrow folksonomies[EB/OL]. [2020－03－25]. http:// www. vanderwal. net/random/entrysel. php? blog = 1635.
[43] ZHOU T C, KING I. Automobile, car and BMW: horizontal and hierarchical approach in social tagging systems [C]// Proceedings of the 2nd ACM Workshop on Social Web Search and Mining. New York: Association for Computing Machinery, 2009: 25－32.

[44] YIN D, XUE Z, HONG L, et al. A probabilistic model for personalized tag prediction[C]// Proceedings of the 16th ACM SIGKDD Conference on knowledge discovery and data mining. New York: Association for Computing Machinery, 2010: 959 – 968.

[45] SIGURBJÖRNSSON B, ZWOL R V. Flickr tag recommendation based on collective knowledge [C]// Proceedings of the 17th international conference on World Wide Web. New York: Association for Computing Machinery, 2008: 327 – 336.

[46] KIPP M E, CAMPBELL D G. Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices [J]. Proceedings of the American society for information science and technology, 2007, 43(1): 1 – 18.

[47] 吴明隆. 问卷统计分析务实——SPSS 操作与应用[M]. 重庆: 重庆大学出版社, 2012: 396 – 397.

作者贡献说明:

庄倩:提出研究问题和研究框架,修改论文;

骆慧颖:研究方案执行,撰写论文初稿;

戴崇丞:数据分析和处理;

刘丽霞:实验数据分析和制图;

靳雪宁:数据收集。

Research on the Influence of User Interaction on the Difference of Social Tagging Behaviors
——A Case Study of Douban. com

Zhuang Qian[1] Luo Huiying[1,2] Dai Dongcheng[1] Liu Lixia[1] Jin Xuening[1]

[1] College of Information Science and Technology, Nanjing Agricultural University, Nanjing 210095

[2] School of Systems Science, Beijing Normal University, Beijing 100875

Abstract: [**Purpose/significance**] From the perspective of users interacting with other users in social tagging systems, the differences in tagging behavior of users with different interaction characteristics are explored. The study is helpful to improve the quality of labels, and optimize the quality of information service of the social tagging systems. [**Method/process**] Using the sample data from Douban Book website, the quantitative indicators were selected to study the distribution characteristics of the users' tagging behaviors from the perspectives of number of tags, tag structure, tag semantics, tagging motivation and user activity. Three indicators including the number of users one follows, the number of one's followers and one's registered age were used to represent a user's degree of association and interaction with other users, then the differences in tagging behavior of users with different interaction characteristics were discussed through difference analysis. Moreover, the influence of the interaction characteristics on these differences were investigated through multivariate regression analysis. [**Result/conclusion**] The results shows that there exists significant differences in social tagging behaviors among users with different interaction characteristics: users who have stronger interactions with other users have more tags; users who have followed more users and have more followers the greater the number of tags used; the longer the user uses Douban, the greater the average tag length and tag reuse, while the relationship with other users has little effect on the average tag length and tag reuse rate; the user's ratio of tags with special language is also greatly affected by the user's registered age, but the number of user's followers does not significantly affect the user's ratio of tags with special language; the more users who follow other users, the more active they are in the tagging system. It is suggested that the social tagging system can take measures to strengthen the interaction between users, and regulate the users' social tagging behavior through the interaction between users, thereby improve the quality of the social tags in the system.

Keywords: user interaction tagging behavior difference